DOCUMENT RESUME

ED 106 309 TM 004 420

AUTHOR Saretsky, Gary
TITLE The John Henry Effect: Potential Confounder of Experimental vs. Control Group Approaches to the Evaluation of Educational Innovations.
PUB DATE 2 Apr 75
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC Not Available from EDRS. PLUS POSTAGE
DESCRIPTORS *Bias; Change Strategies; *Control Groups; *Educational Innovation; Evaluation Methods; *Experimental Groups; Experiments; Innovation; Literature Reviews; Organizational Change; Performance Factors; Research; Research Design; Research Methodology; *Research Problems; Technological Advancement
IDENTIFIERS *John Henry Effect

ABSTRACT

The paper describes an unacknowledged artifact that may confound experimental evaluations of innovations. The paper hypothesizes that control group members (teachers, pupils, etc.) perceiving the consequences of an innovation as threatening to their job, salary, status, or traditional patterns of working, may perform atypically and confound the evaluation outcomes. Demand characteristics within the social psychology of the experiment provide the theoretical framework. The paper compares similarities and differences among the John Henry Effect and other research-biasing factors. Four evaluation studies in which the John Henry Effect was manifested are described. Alternative evaluation designs for the artifact's control are discussed. (Author)

ED106309

THE JOHN HENRY EFFECT: POTENTIAL CONFOUNDER OF EXPERIMENTAL vs CONTROL GROUP APPROACHES TO THE EVALUATION OF EDUCATIONAL INNOVATIONS

Gary Saretsky

Center for Evaluation, Research and Development

Phi Delta Kappa International

The paper describes an unacknowledged artifact that may confound experimental evaluations of innovations. The paper hypothesizes that control group members (teachers, pupils, etc.) perceiving the consequences of an innovation as threatening to their job, salary, status, or traditional patterns of working, may perform atypically and confound the evaluation outcomes. Demand characteristics within the social psychology of the experiment provide the theoretical framework. The paper compares similarities and differences among the John Henry Effect and other research-biasing factors. Four evaluation studies in which the John Henry Effect was manifested are described. Alternative evaluation designs for the artifact's control are discussed.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS [redacted] MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY
GARY SARETSKY
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

TM 004 420

Presented at the American Educational Research Association's Annual Meeting, April 2, 1975, Washington, D.C.

Most of you, no doubt, are familiar with the American folk hero John Henry, and the ballad of John Henry. John Henry, as you recall, was a 19th Century rail driver, who swung his 16 pound hammer driving spikes and drill bits. The "Ballad of John Henry" tells of his competition with a steam drill, an innovation which eventually replaced rail drivers. All day and all night the competition went on. In the end John Henry outperformed the steam drill. But this triumph was bittersweet and short-lived for John Henry died the next day from the overexertion of the competition.

This folk ballad has some interesting implications for those of us who would apply classical experimental designs to the evaluation of technological innovations in education. If we were to cast this tale into an experimental mode we would probably label the steam drill the experimental treatment and John Henry, using his 16 pound hammer, as the control treatment. Most critiques of such an experiment would emphasize the small n, the selection procedure as a source of bias, and perhaps, due to John Henry's demise, the non-replicable nature of the experiment. The folk ballad, however, highlights a far more significant biasing factor, that of the extraordinary, atypical effort of those executing the control treatment. I suggest that if you were to examine most large scale evaluations of technological innovations e.g., evaluations of instructional television or computer assisted instruction (CAI), you would find little, if any, evidence of attempts to ascertain the "normalcy" of control group behavior. Under such designs any atypical performance of those executing the control, or more appropriately the comparison treatment, would likely go undetected,

confounding the results of the evaluation and thereby fundamentally misleading educational decision makers regarding the substantive worth of the innovation.

Could John Henry's extrordinary performance due to his perception of the consequences of the innovations superior performance as threatening to his status, job, salary, or traditional patterns of work? It is difficult for us to know. But what I'd like to suggest in this presentation is that the perception of such threats are characteristic of reactions to the introduction of highly technological innovations in education, furthermore, that these reactions often stimulate atypical performance by those representing the status quo, the very group that typically constitutes the control group in experimental approaches to the evaluation of innovations, and lastly, that this source of bias and the consequent biased responses (what I refer to as the John Henry Effect) have led to many of the N.S.D. (non-significant difference) findings that have characterized so much of evaluation research.

In the remaining time I would like to 1) briefly describe the constructs that would be supportive of the John Henry Effects potential manifestation, 2) distinguish the John Henry Effect from other acknowledged research biasing factors and 3) describe four evaluation case studies which would be illustrative of instances where the John Henry Effect should be considered as an alternative explanation of the evaluation's outcomes.

There are two areas of inquiry germane to the hypothesized artifact. The first relates to studies of receptivity and resistance to technological change and social innovation. The second area relates to studies of the social psychology of the experiment (Orne 1962, 1969) and more particularly to the artifacts (research biasing factors) that arise therin.

Resistance to Change and Innovation

As early as the fifteenth century, Niccolo Machiavelli wrote of the difficulty of introducing change and innovation because of possible reactions. He commented:

> It must be considered that there is nothing more difficult to carry out, nor more doubtful of success, nor more dangerous to handle, than to initiate a new order of things. For the reformer has enemies in all those who profit by the old order, and only lukewarm defenders in all those who would profit by the new order, this lukewarmness arising partly from fear of their adversaries, who have the laws in their favor; and partly from the incredulity of mankind, who do not truly believe in anything new until they have had actual experience of it. Thus it arises that on every opportunity for attacking the reformer, his opponents do so with the zeal of partisans, the others only defend him halfheartedly, so that between them he run great danger. (Niccolo Machiavelli, The Prince)

More recently, in reference to the problem of change in the school culture, Sarason observed:

> It will be, I think, axiomatic in a theory of change that the introduction of important change does not and cannot have the same significance for the different groups comprising the setting, and that one consequence is that there will be groups that will feel obligated to obstruct, divert or defeat the proposed change. (Sarason, 1971)

It is therefore not surprising to find that many attempts at innovation fail. A number of factors relevant to the hypothesized artifact operate in the organizational innovative process and may motivate organization members, either conciously or unconciously to obstruct, divert or defeat the proposed changed. As indicated by Webb et al. (1966) classical approaches to evaluation, using experimental methodology, are often insensative to causal factors and fail to differentiate the effects resulting from such factor's manifestations and those of the innovation or "treatment."

In identifying factors that may cause resistance, Havelock (1969) indicates that there is a need for stability within organizations. Because change is disruptive, it is likely to be resisted. One problem is an innovation's potential impact on existing social relationships within the organization (Stewart, 1957). This source of resistance has been discussed by Schon (1967), who says, "Innovation threatens also the hierarchy of power and prestige on which the corporation's system of control is built, for its political structure is tied to an established technology." Since the power and prestige hierarchy may be vulnerable to disruption, change can appear to be a personal threat. For example, the idea of hiring consultants may produce objections based on a perceived danger to existing roles and prestige structures; those who resist the entry of a consultant may be worried that he or she will criticize their role performance or suggest that their role specifications be altered (Zaltman et al., in press).

Status problems can also arise in the form of a status discrepancy between the recipient and donors of new knowledge (Rice, 1963). When the donor organization has high status relative to the recipient organization, there will be barriers to the information flow between them (Czepiel, 1972). Researchers have explained that by seeking or accepting new information, organization members seem to be admitting inferiority. In an analagous fashion, classroom teachers often dismiss the advice of university professors using the excuse that the professor doesn't understand the problems of the public school practitioner.

Classroom teachers may also be motivated by local pride, a source of resistance discussed by Havelock (1969). When local personnel believe in the unique and positive qualities of their organization, they mistrust new knowledge as potentially harmful. Scientists in organizations (Allen cited by Havelock) and administrators in business firms (President's Conference, cited by Havelock) have been shown to exhibit such resistance.

After the initial entry of new information or knowledge, the stages of attitude formation and decision making also contain the seeds of resistance to innovation. As Havelock (1969) points out, "Internally the organization can be seen as a complex system of filters; each subsystem and each member has some power to block the flow of information, to screen it, censor it, and distort it." Watson (1973) discusses an associated factor, which he terms systemic coherence, noting that a change in one part of a system must affect other parts. An example cited by Watson is a technological change which added so much to the productivity of piece workers that they earned more money than their supervisors; this innovation was then rejected.

An analogous problem occurred during the recent Office of Economic Opportunity's experiment in performance contracting, where paraprofessionals, using special materials and a token reinforcement system, were to receive bonuses based on student test score gains. As discussed in a preliminary study of the John Henry Effect, this "piece work" was cited as unprofessional and contributed to the demise of the innovation(Saretsky, 1972).

Havelock (1969) suggests that roles have the effect of inhibiting innovation and preserving the status quo: "Most role expectations are

designed to stablize and routinize human performance. They encourage conformity. . .The more sharply defined and the more limited the role, the less room there will be for receiving and sending messages which are 'new' and hence different from what is expected."

An organization's reward patterns may be structured so as to favor conservative rather than innovative decision making, when staff members are rewarded for stable, reliable behavior (Rother 1960, Schon 1967). The salary and promotion structures of most educational institutions reinforce and reward longevity and doing "more of the same." Educational systems usually lack the mechanisms to reward innovative behavior.

Resistance can arise and be effective at the stage of implementing an innovation. There are various strategies that an organization may use to deal with an innovation that is perceived as threatening. As Graziano (1969) indicates:

> It might incorporate the new event and alter it to fit the preexisting structure so that, in effect, nothing is really changed. It might deal with it also by active rejection, calling upon all of its resources to 'starve out' the innovator by insuring a lack of support.
>
> The most subtle defense, however, is to ostensibly accept and encourage the innovator, to publicly proclaim support of innovative goals, and while doing that to build in various controlling safeguards, such as special committees, thereby insuring that the work is always accomplished through power structure channels and thus effecting no real change.
>
> ...Innovation is thus allowed, and even encouraged, as long as it remains on the level of conceptual abstractions, and provided that it does not, in reality, change anything!

Individuals may also manifest passive resistance as a strategy for rejecting innovation. They may simply fail to follow the instructions of

management, or they may implement the innovation in a partial or dysfunctional way. Zaltman et al. (in press) gives the example of educational simulation games, which were ordered by principals for use in the classroom. When the games were left on the shelf, or used halfheartedly and thus somewhat ineffectively, some school principals decided to abandon the concept of academic gaming. This was an instance in which an innovation was supposedly given an informal trial, and in which lack of cooperation on the part of teachers resulted in disappointing results. In this case, the teachers who were asked to use the innovation failed to do so; in other cases discussed below, when a control group of teachers was set up, these teachers performed atypically.

Much of the inquiry into reactions and receptivity to technological and social change posts an analytical framework of change as a multi-phased reaction process, a process in which final outcomes will be the resolution of mitigating interpretations and accomodations (King and Repton, 1968). Figure I represents one such process.

With the exception of stage one in Figure I, each subsequent stage (e.g., stages 2, 3, and 4) are entered into only if warranted by a reaction similar to that described in the extreme right hand column. If, for instance, the innovation is perceived as having no threat to the present status, security and work definition of a homogeneous occupational group (i.e., classroom teachers) social barriers to the implementation of the change are not produced*. If, however, threats are perceived, stage two is entered into where these threats are confronted, and the inherent ability

*These social barriers are distinguished from structural barriers (i.e., legal, fiscal or technological, etc.), impediments that may yet exist.

Figure I. A MULTIPHASED REACTION-TO-CHANGE-AND-INNOVATION PROCESS

Phase	Description	Pivotal factors	Reaction within stage	Reaction leading to sub-subsequent change
I	Interpretation and projection of possible consequences	considerations of econmic security occupational security, occupational prestige, work	Change is interpreted as having no threat to present status, security and work definition	Change is interpreted as having threat to present status, security and work definition
II	Confrontation of change	degree of stress (none to intense) resulting	flexible absorbtion of modicum of change	generalized anxiety, dissatisfaction, and frustration
III	System adjustment	flexibility of social system	accomodation, over period of time, by stretching mechanisms that sustain the basic social system	disorganization of basic mechanisms that sustain the basic social system
IV	System reorganization	modification of system	accomodation of stresses produced by changes, curtailment of disorganizing consequences threatening survival of the system	

of the social system to acclimate to a modicum of change is tested. If the change is perceived as major, or if the inherent flexibility of the system is insufficient, the more advanced stages are entered into leading either to accomodation over a longer period of time, or to disorganization and a subsequent reorganization of the entire system.

As indicated in the preceding review of the literature, change and innovation presents, or can be perceived as presenting, threats to the jobs, salaries, status and the traditional working patterns of certain individuals representative of the status quo. Furthermore, the review indicates that reactions to such perceived threats take the form of actions designed to accomodate, control, thwart or defeat the proposed change and/or innovation. This review continues with a description and comparison of research biasing factors currently considered as rival hypotheses when experimental outcomes are inconsistent with prevailing laws, theories and logical expectations or are counterintuitive to those involved in the conceptualization and implementation of the innovation's experimental evaluation.

Research Biasing Factors as Alternative Explanations

Calling attention to the social psychology of the experiment, Orne (1962) observes that much of human behavioral research focuses upon what is done to the subject rather than what the subject does in reaction to the cues and stimuli of an experiment. The former category--what is done to the subject--has been the focus of most inquiries into research biasing factors in education, including research on the Placebo Effect, the Experimenter Bias Effect, the Investigator Bias Effect, the Halo Effect, and the Hawthorne Effect. Interestingly, much less educational research has been directed toward the latter category--what the subject does. Clinical psychology, however, has identified three research biasing factors, Demand Characteristics (Orne 1962), the Deutero Problem (Reicken

1962), and Evaluation Apprehension (Rosenberg 1969). A brief description of these factors and a comparison of their attributes are displayed in the facet analysis are displayed in the facet analysis (Figure II) on page 18.

The Halo Effect refers to the tendency, in making an estimate or rating of one chracteristic of a person, to be influenced by another characteristic or by one's general impression of that person (Medley and Mitzel, 1963). The Halo Effect manifests itself most commonly in the rating of a person's performance or a product of that person's performance. Kerlinger (1964) offers as examples, "The professor assessing the quality of essay test questions higher than they should be because he likes the testee. Or the rating of the second, third and fourth questions higher (or lower) than they should be because the first question was well (or poorly) answered."

The Placebo Effect has it origins in biomedical, pharmacological, and psychopharmacological research. It refers to the therapeutic effect that a chemically inert substitute (such as sugar) has upon the patient when the patient (and doctor), unaware of the substitution, believe in the efficacy of the medication. In social service programs an attempt is occasionally made to control for this effect by setting up". . .equally elegant appearing treatments . . ." to be given to two groups (Anderson et al., 1974), one treatment being the innovation, the other a placebo or substitute which by itself should not have an effect. Suchman (1967), however, states that the notion of setting up such "sunmy" programs and utilizing double blind designs (where patient and doctor, or student and

teacher, are unaware of the substitution) is usually impractical for the evaluation of complex social innovations.

Experimenter Bias Effect refers to an experimenter's unintentional and unconscious communication of his or her expectancies of the experimental outcome as a partial determinant of those outcomes (Rosenthal 1963, Barber 1973). This subtle outcome bias alters the normal functioning of the subject on the dependent variable(s) central to the research study. Examples of the subtle communications are unintentional verbal and visual reinforcement--i.e., smiles, grunts, nods--of responses consistent with the hypothesis.

In discussing the Investigator Bias Effect, Barber (1973) distinguishes between the role of the investigator, who is the conceptualizer and designer of the research activity, and the role of the experimenter, the individual(s) who interacts with the subject, administers the treatment, and makes observations. As indicated in the procedures section of proposal, Barber contends that the paradigm within which the investigator works determines the nature of the hypothesis, the variables selected, the data deemed relevant, and the subsequent analysis and interpretation of the results. Such Investigator Bias would inhibit the investigator's consideration of alternative hypotheses, designs, and interpretations, i.e., considering the John Henry Effect as an artifact which would confound the Experimental vs. Control evaluation outcomes, and recognizing the necessity for alternative designs or procedures to control for such an artifact.

the necessity for alternative designs or procedures to control for such an artifact.

The Hawthorne Effect refers to unanticipated but beneficial effects produced in experimental situations. Such effects are said to be caused by the subject's awareness that he or she is in an experiment and the object of special attention (whether real or imagined), an awareness that is said to have a positive effect on the subject's performance during the duration of the experimental period (Cook, 1967).

Unfortunately, the Hawthorne Effect has been used as a general rubric under which researchers "have swept unexpected, striking results which defied explanation in line with the procedures used and pre-existing information" (Gephart and Antonopolos, 1969). Most standard texts on educational research methodology warn researchers to beware of the effect, yet they provide only vague descriptions of the phenomenon and its impact upon experimental designs, and only vague suggestions for its control (Cook, 1967).

The most ambitious and systematic attempt to study, define, operationalize, and control for the Hawthorne Effect was headed by Cook (1962, 1963, 1967). Through a systematic and exhaustive search of the literature, the inquiry found many inconsistent and contradictory references to the Hawthorne Effect, which was variously attributed to:

1. Novelty - as in the novelty of a new experimental technique, i.e., computer assisted instruction, or a different experimental setting.

2. Awareness of Participation - the subject perceives himself as a guinea pig and object of experimentation.

3. Altered Social Structure - as in an experimental situation where management increases deference to the subjects and allows subjects to participate in local decision making.

4. Knowledge of Results - it is suggested that informing the subject of his rate of productivity will be reinforcing and provide the subject a level of self-performance to compete with.

As a working definition, Cook described the Hawthorne Effect as:

> ...a phenomenon characterized by a cognitive awareness on the part of the subjects of special treatment created by artificial experimental conditions. It becomes confounded with the independent variable under study with the subsequent result of either facilitating or inhibiting the dependent variables under study and leading to spurious conclusions. (Cook 1962, 1967).

Such an all-encompassing "working" definition was necessary considering the inadequate and often contradictory literature and research evidence of the Hawthorne Effect.

Interestingly, Cook's two-year experimental study of the variables which various writers had cited as component factors of the Hawthorne Effect (1967) failed to reveal evidence of a Hawthorne Effect. Cook concluded that:

> ...it appears unlikely that one can employ a Hawthorne Effect concept to explain differences or the lack of differences between experimental and control groups in educational research studies insofar as the variables commonly believed to generate the effect such as direct and indirect cues, the duration of a study, and mechanical changes introduced in an experiment are considered to be sufficient potency to produce the effect. (1967)

Two more experimental studies of the Hawthorne Effect (Rubeck 1971, Bauernfeind and Olson 1973) were also unable to experimentally induce the Hawthorne Effect and came to similar conclusions, namely, that their findings:

> ...raised major doubts about the Hawthorne Effect as a confounder of educational experimental results. In short, it appears that either the Hawthorne Effect as presently conceived, or the present study, is open to serious question.
>
> Despite the great number of conjectural writings that caution us to protect our experiments from the Hawthorne Effect (or "reactive effects"), the only empirical studies of such possible effects--the present study--(Rubeck's study) and Cook's study have failed to disclose them.
> (Bauernfeind and Olson, 1973)

Demand Characteristics refer to Orne's (1962) hypothesis that each experiment creates demands on the subject that are of the subject's own making. The subject's knowledge that he is in an experiment cause him to try to ". . . ascertain the true purpose of the experiment . . ." so as to respond in an appropriate manner. The subject searches for cues which will indicate what the hypothesis is:

> The totality of cues which convey an experimental hypothesis to the subject become significant determinates of the subject's behavior.
>
> These cues include the rumors or campus scuttlebutt about the research, the information conveyed during the original solicitation...the experimenter...the setting...all explicit and implicit communications during the experiment proper... (and) the experimental procedure itself viewed in the light of the subject's previous knowledge and experience (Orne 1962, Orne and Holland 1972).

The hypothesis perceived by the subject may be totally different from the experimenter's, for it is dependent upon the particular combination of cues and interpretation which the subject selects.

The Deutero Problem refers to the dilemma or problem that a subject is unconsciously faced with when he must chose between being a "good subject" and winning the experimenter's approval, and meeting personal needs, e.g., the need to succeed, the need to protect himself (Reicken 1962). The effort to address these needs may be a significant determinant of the subject's performance.

Evaluation Apprehension refers to an experimental subject's anxiety-toned concern that he win a positive evaluation from the experimenter, or at least that he provide no grounds for a negative one (Rosenberg, 1969). An individual being evaluated would therefore perform atypically. This behavior is not only evident in clinical settings, but can be observed in the classroom when the teacher is evaluated by his supervisor or principal.

The John Henry Effect's existence was first suggested by Robert Heinich. Explaining the difficulty experienced by advocates of mediated forms of instruction in demonstrating the superiority of their innovations, he commented:

> One of the reasons why no statistically significant differences conclusions result from so many television versus classroom teaching experiments may be that classroom teachers are spurred to "maximum" performance, a condition I have referred to as the "John Henry effect." Evidence of this was indicated in the Anaheim, California, television experiment, by Dr. Kenneth D. Hopkins, one of the principal investigators, in a public statement at the University of California. Each successive year of the five-year experiment witnessed a drop in classroom teacher performance while the mediated instructor remained the same. (However, all indications are that classroom teaching is considerably improved as a result of television teaching, and remains above prior levels.) Looking at this another way, if televised teaching had been measured against classroom teaching of the year before the experiment began, the results might have been quite different: Or if it at all possible, an experiment should be conducted where the classroom teachers in a district are unaware that a comparison is being made so that typical performance is measured; again the results might be quite different. (Heinich 1970)

Subsequent inquiry (Saretsky, 1972, 1972a) has delineated the John Henry Effect as the confounding influence that undetected atypical performance, aroused by perceptions of an innovation's threat to jobs, status and work patterns, has upon an experimental evaluation of that innovation. Such perceived threats are associated with innovations that a) substantially alter the roles and relationships within an occupational setting e.g., the delivery of instruction transferred from a certificated teacher to an interactive computer based delivery system, b) replaces the worker with individuals of lesser status e.g., replacement of certificated teacher by paraprofessional and a modular instructional system, or c) an innovation that threatens the worker's salary e.g., payment of a teacher on a piece work basis, or basing teacher salary upon student performance.

The context in which this atypical behavior would arise would be one of stress, e.g., where accountability systems were being installed, or where administrative pressures for economies and efficiencies were evident.

Despite similarities among the research biasing factors listed above, they do show differences on at least two dimensions: the locus of the effect and the effect and the nature of the error contributed. These factors can be displayed (after Gephart and Antonoplos, 1969) on three facets of comparison. These facets are:

1. The central aspect of the concept. Each of these biases appears to have a focus of operation, a place in which the activities which create them congeal into the effect.

2. The location within the research process. The concepts deal with aspects of the research process which if expressed on a time dimension display differentiation.

3. The kinds of error contributed. Common to all of these concepts is the idea of a contribution to a conclusion on the part of the researcher that diverges from truth in an absolute sense. There are many possible kinds of error that can be identified in the research process. Again, the concepts differ.

The grid (see Figure V, page 30) indicates the levels on these three facets which depict the different biasing concepts. The reader will note that several definitions for the Hawthorne Effect are given, in accord with Cook's work.

The specific nature of the factors contributing to these artifacts' manifestations, their interaction with the type of innovation and the experimental setting, and the magnitude of their impact upon decisions in different decision settings are yet to be determined and will be the focus of further inquiry.

FACETS FOR COMPARISON			
	Central Aspects	*Location Within the Research Process*	*Kinds of Error Contributed*
EXPERIMENTER BIAS EFFECT	Expectancies held by experimenter and their effect on his behavior with subjects.	Structuring of procedures and experimenter-subject interaction.	Modification of the treatment with subsequent threat to the internal validity of the test of the hypothesis.
HAWTHORNE EFFECT			
1. Novelty	Interaction between subject and research procedures.	Initial interaction between subject and procedures.	Modification of the treatment with subsequent threat to the internal validity of the test of the hypothesis.
2. Awareness of Participation	Same	Throughout the research process.	Same as above.
3. Altered Social Structure	Interaction between subject, other subjects, and experimenter.	Interactions between individuals	Same as above.
4. Knowledge of Results	Interaction between subject and a specific aspect of the research procedure.	Follows reporting of subject's performance.	Same as above.
DEMAND CHARACTERISTICS	Subject's perception of his role in the experiment.	Continuous	Modification of the subject's role with subsequent threat to the external validity of the test of the hypothesis.
HALO EFFECT	Rater's reaction to non-relevant information in rating process.	During measurement involving ratings.	Measurement error not necessarily common across subjects.
PLACEBO EFFECT	Control subject's interaction with research procedures.	During experimental and control procedures.	Alters performance of control subjects, resulting in an inaccurate comparison between groups.
INVESTIGATOR BIAS EFFECT	Paradigim under which investigator designs, carries out, and interprets research	Design of experiment, generation of hypothesis, selection of variables, subjects, and analysis procedures, and analysis and interpretation of outcomes.	Modification of factors with resultant threats to internal and external validity of the test of the hypothesis.
DEUTERO PROBLEM	Choice between being "good subject" and meeting personal needs.	Initial interaction between subject and experimenter.	Alters performance of subjects with subsequent threat to external validity of the hypothesis.
EVALUATION APPREHENSION	Subject's anxiety of evaluation and subsequent behavior to avoid negative evaluation.	Initial interaction between subject, experimenter and procedures.	Alters performance of subjects.
JOHN HENRY EFFECT	Subject's perception of consequences of innovation and subsequent behavior to demonstrate superiority of traditional methods or avoid negative evaluation or to retain status and traditional patterns of work.	Interaction between subject, experimenter and procedures.	Modification of subject's performance with subsequent threat to internal validity of the hypothesis.

FIGURE II

Facet Comparison of Artifacts

The facet comparison (Figure II) yields the most interesting comparisons among the Hawthorne Effect (HE), the Evaluation Apprehension Effect (EA), and the John Henry Effect (JHE). The major distinctions between the John Henry Effect and the variations of Hawthorne Effect are the focus of the former (JHE) upon consequences and perceived threat, and the focus of the latter (HE) upon the initial awareness or interaction within the process, and its general association with such terms as enthusiasm, or facilitative effects. Although not mutually exclusive classifications the Hawthorne Effect is also usually associated with the experimental treatment, whereas the John Henry Effect is usually associated with the control (or comparison) treatment.

Evaluation Apprehension shares with the John Henry Effect, the antecedent of perceived threat. The threat in the former (EA) is associated with the process of being evaluated, whereas the threat in the latter (JHE) is associated with the consequences of the innovation and its effects upon jobs, status, salary and traditional work patterns. Research associated with Evaluation Apprehension have found more significant effects with responses to evaluations of psychological or social deviancy and fewer significant effects with evaluations of work performance. It is with the evaluations of work performance and productivity that the John Henry Effect is associated.

It has been hypothesized in a preliminary study (Saretsky 1972a) that were a John Henry Effect present, it could be displayed in one or more of the following manners (see Figure IV, page 21)

Figure IV a represents the control group markedly outperforming the experimental group.

Figure IV b represents the control group outperforming both the experimental group and the control group.

Figure IV c represents the discrepancy between the predicted performance of a control group and their actual performance.

Figure IV d represents the variation in control group performance prior to, during, and after the experimental evaluation.

FOUR CASE STUDIES

The four evaluation case studies described in the next section of this paper are exemplars of evaluations in which the manifestion of the John Henry Effect should be considered as an alternative explanation for the effects observed.

---

1. Zdep and Irvine (1970) described an experiment designed to "assess the effect of supportive radio and television broadcasting of English instruction among fifth grade sutdents in northern Nigeria." At the conclusion of the study, but prior to the data analysis, the researchers were told by the school's headmistress, ". . . that the teacher of the control class had periodically expressed her displeasure at not having been selected to teach one of the classes having supplemental television or radio broadcasts."

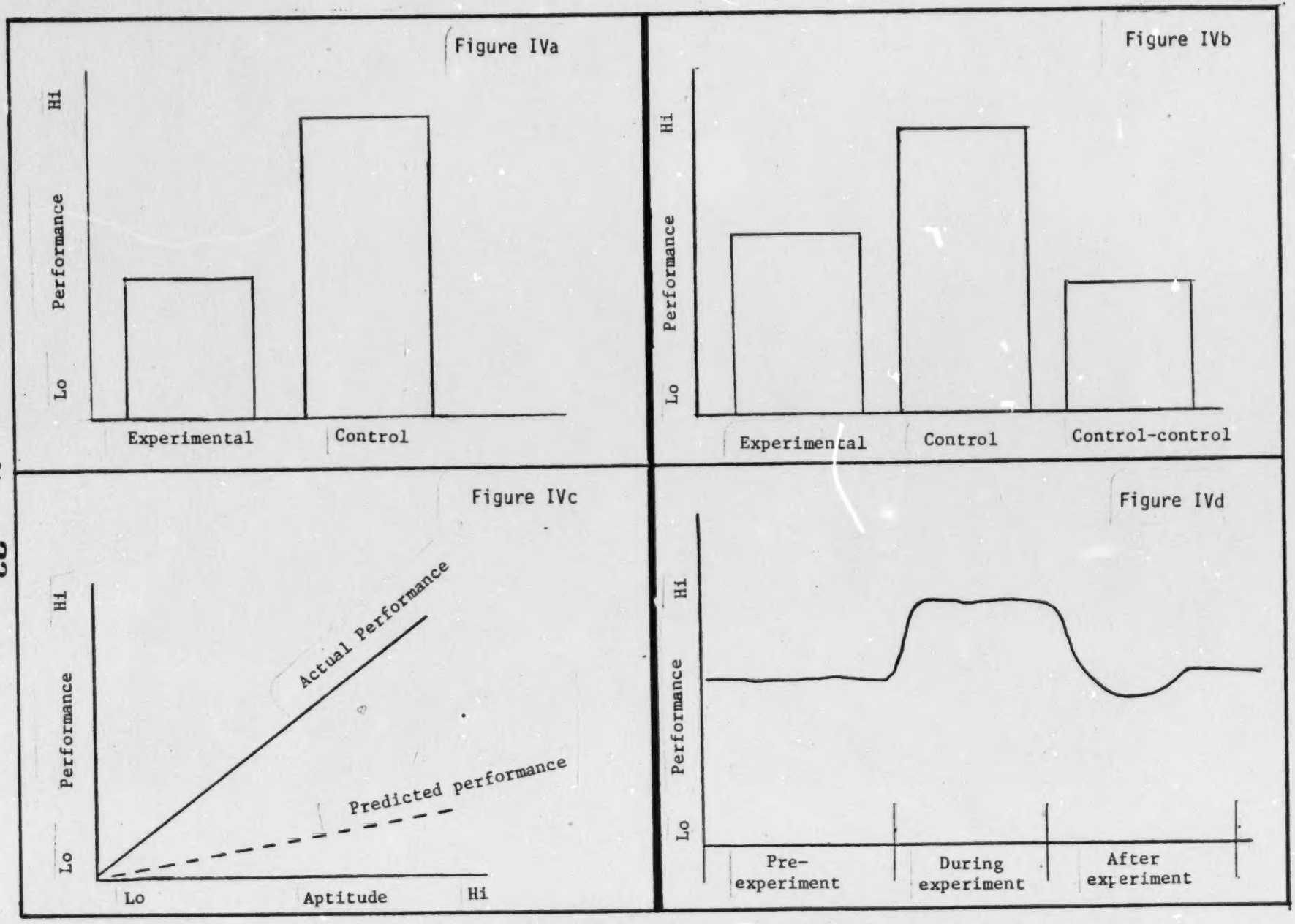

Figure IV Graphic configurations of The John Henry Effect

The researchers were unable to communicate the idea of random assignment and the teacher therefore felt slighted by her designation as a control teacher. "This teacher further stated that she would do everything possible to have her class do better in English than the experimental classes."

The results of the data analysis revealed that the control teacher's class out-performed the experimental classes. Zdep and Irvine concluded that control groups in certain educational evaluations may not provide the bias-free base lines deemed necessary for comparison, especially when experimental and control groups are housed at the same school.

2. Pella, Stanley, Wedemeyer, and Wittich (1962) reported a study which compared physics instruction supplemented by the Harvey White Physics film series (T-films), a control group receiving conventional instruction (C) and a third control group (CC) whose teachers and pupils had not seen the films and were in no way associated with the project until seven days before the posttest. The second control group (CC) was designed to control for a possible Hawthorne Effect (increased enthusiasm and effort due to a group's knowing that it is participating in an experimental situation).

The control Group (C) students outperformed both the groups receiving supplemental film instruction (T-film) and the second control group (CC). The superior performance of the control group (C) was "assumed" to be the result of the Hawthorne Effect (1962). However, this interpretation appeared to be contrary to the findings of Cook (1967), Rubeck (1971), and Bauernfeind and Olson (1973).

In subsequent conversations with the researchers in this study (Saretsky 1972a), it was revealed that the teachers resented being replaced

by films. Stanley stated that control group (C) teachers put an extra effort into their teaching and devoted extra time to developing experiments and presentations. They knew that they were being evaluated and wanted to look good in comparison to the films. Furthermore, Stanley pointed out that because of the random assignment procedures, physics teachers with negative attitudes toward the films were sometimes selected as experimental (T-film) teachers. They intentionally performed poorly in class, just sitting and doing nothing other than showing the films--which were supposedly only supplemental. "They weren't going to be shown up by any films," was the way Stanley put it. (Saretsky 1972a)

3. Suppes (1969) reported a comparison of a computer-assisted instruction program in mathematics with conventional instruction. Two of the control groups performed significantly better than those receiving computer-assisted instruction in mathematics. The differences were so striking that Suppes examined in detail what the control treatment was. As it turned out, immediately upon being designated as control group teachers, the control teachers went out and purchased additional drill and practice workbooks for the control students. No explanation for the teachers' behavior was provided, but in a subsequent conversation (Saretsky 1972a) Suppes hypothesized that the teachers wanted to demonstrate that they could do as well or better than the computer. Unfortunately, this was not the hypothesis to be tested by the study.

4. In an analysis of the Office of Economic Opportunity's experiment in performance contracting, Saretsky (1972) observed some unusual gains made by control group students. These gains of up to 1.6 years in reading and math were made by students with a history of poor achievement.

In addition to identifying selection and regression effects as partial determinants, the study obtained anecdotal evidence from project directors, site evaluators, the management support group, and teacher union representatives to the effect that control group teachers were performing atypically during the experiment. These interviews evoked comments such as, "When you entered the control school, you knew the race was on," "Those teachers were out to show that they could do a better job than those outsiders (performance contractors)," and, "I don't have any hard data on it because we weren't required to get it, but I know those teachers just worked harder." In its Report to Congress (1973), the Office of Management and Budget also questioned the "no significant differences" conclusion of OEO, citing evidence of atypical performance by control group teachers.

In each of these last three Evaluation versus Control studies, the innovation being evaluated would result in alteration of the role and traditional pattern work of the control group. In each of these studies, there was evidence of atypical control group performance which confounded the results of the study, thereby nullifying the credibility and utility for educational decision making of the information derived.

SUMMARY

The preceding paper posts that when a technological innovation is introduced in certain social settings the consequences of that innovation may be perceived as threatening to the jobs, status, salary and/or traditional work patterns of those who constitute the status quo. Research on reaction and resistance to change and innovation suggest a number of reactions that those representing the status quo may make. Atypical performance is one class of such reactions. When the introduction of a highly technological innovation is comingled with an experimental evaluation of that innovation, the aroused atypical performance of a control group comprised of those who represents the status quo may go undetected, thereby confounding the evaluation, and misleading decision makers as to the substantive worth of the innovation. Although sharing attributes with other commonly recognized artifacts, a facet analysis reveals unique characteristics of the artifact described in this paper as the John Henry Effect.

McGuire (1969) depicts three stages in the life of an artifact: a) the ignorance stage, b) the stage of coping, and c) the exploitation stage. In the first stage researchers and evaluators seem unaware of the variable producing the artifact and tend to even deny it when its possibility is pointed out to them. The second stage begins as its existance and possible importance become undeniable. In this coping phase, researchers then begin to recognize and even over-stress the artifact's importance. They give a great deal of attention to devising procedures which will reduce the artifact's contaminating influence and its limiting of the generalizability of experimental results. Evaluators pursue similar actions to insure the

validity of information provided to decision-makers. The third stage, exploitation, grows out of the considerable intellectual effort during the coping stage to understand the artifactual variable so as to eliminate it from the experimental situation. In their attempt to cope, some researchers almost inevitably become interested in the artifactual variable in its own right. It then begins to receive research attention, not as a contaminating factor to be eliminated, but as an interesting independent variable in its own right.

The purpose of this presentation was to make you aware of the John Henry Effects possible existence and to stimulate your interest in initiating a programmatic effort leading from stage one to stage two in the life of the John Henry Effect.

REFERENCES

Allen, W.H. Instructional media research, past, present and future. AV Communications Review, Spring, 1971, 19, 1, pp. 5-18.

Anderson, S.B., Ball, S., Murphy, R.T., and Rosenthal, E.J. Anatomy of Evaluation: Important Concepts and Techniques in Evaluating Education/Training Programs. Educational Testing Service. Princeton, New Jersey, 1973.

Barber, T.X. Pitfalls in research: nine investigator and experimenter effects. In R. Travers (Ed.). Second Handbook of Research on Teaching. Rand McNally and Co. Chicago, Illinois, 1973.

Bauernfeind, R.H. and Olson, C.J. Is the Hawthorne effect in educational experiments a chimera? Phi Delta Kappan, 1973, 55, pp. 271-273.

Campbell, D.T. Conversational contact. Evaluation, 1973, 1, pp. 7-13.

Campbell, D.T. and Stanley, J.C. Experimental and Quasi-experimental Designs for Research on Teaching. Rand McNally and Co. Chicago, Illinois, 1963.

Carrese, L.M. and Baker, C.G. "The Convergence Technique: A Method for the Planning and Programming of Research Efforts," Management Science, Vol. 13, No. 8, April 1967.

Cook, D.L. The Hawthorne effect in educational research. Phi Delta Kappan, 1962, 44, pp. 116-122.

Cook, D.L. The Impact of the Hawthorne Effect in Experimental Designs in Educational Research. United States Office of Education, Cooperative Research Project, Washington, D.C., 1967, No. 1757.

Cook, D.L. The relation of possible Hawthorne effect components to stages of experimental investigation. Occasional papers, Ohio State University, 1963.

Czepiel, J.A. The diffusion of major technological innovation in a complex industrial community: an analysis of social processes in the American steel industry. Ph.D. dissertation, Northwestern University, 1972.

Gephart, W.S., Ingle, R. and Saretsky, G. Similarities and Differences Between Research and Evaluation. Phi Delta Kappa, Bloomington, Indiana, 1973.

Gephart, W.S. Application of the Convergence Technique to Basic Studies of the Reading Process. Final Report, U.S. Department of Health, Education and Welfare, Office of Education, 1970 ED-037-587.

Gephart, W.S. and Antonoplos, D.P. The effects of expectancy and other research biasing factors. Phi Delta Kappan, 1969, 50, pp. 579-583.

Graziano, A.M. Clinical innovation and the mental health power structure: a social case history. American Psychologist, 1969, 24, pp. 10-18.

Guttman, Louis. A structural theory for intergroup beliefs and actions. American Sociological Review, 1959, 24, 318-328.

Hage, J. and Aiken, M. Social Change in Complex Organizations. Random House. New York, 1970.

Havelock, R.G. Planning for Innovation through Dissemination and Utilization of Knowledge. Center for Research on Utilization of Scientific Knowledge, University of Michigan. Ann Arbor, 1969.

Heinich, R. Technology and the Management of Instruction. Department of Audio-Visual Instruction, Inc. Washington, D.C., 1970.

Kerlinger, F.N. Foundations of Behavioral Research. Holt, Rinehart and Winston, Inc. New York, 1964.

Klein, S. Evaluation of the Office of Economic Opportunity Performance Contracting Experiment. Report to Congress by the Comptroller General of the United States, 1973.

Kling, M. Developing a data base for model building in reading. Twentieth Yearbook of the National Reading Conference. The National Reading Conference, Milwaukee, 1970.

Kuhn, T.S. The Structure of Scientific Revolutions. Chicago, University of Chicago Press, 1962.

Machiavelli, N. The Prince. Modern Library. New York, 1967.

Medley, D.M. and Mitzel, H.E. Measuring classroom behavior by systematic observation. In N.L. Gage (Ed.), Handbook of Research on Teaching. Rand McNally and Co. Chicago, 1963.

Orne, M.T. On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications, American Psychologist, 1962, 17, pp. 776-93.

Orne, M.T. and Holland, C.H. On the ecological validity of laboratory deceptions. In a. Miller (Ed.), The Social Psychology of Psychological Research. The Free Press. New York, 1972.

Pella, M.O., Stanley, J.C., Wedmeyer, C.A. and Wittich, W.A. The use of the White films in teaching of physics. Science Education, 1962, 46, 1, pp. 6-21.

Reicken, H. A program for research or experiments in social psychology, Decisions, Values and Groups. Washburne, N.F. (Ed.) Pergamon Press, New York, 1962.

Rice, A. The Organization and Its Environment. Tavistock Publications. London, 1963.

Rosenberg, M.J. The conditions and consequences of evaluation apprehension. Artifacts in Behavioral Research, Rosenthal, R. and Rosnow, R. (Eds.) Academic Press, New York, 1969, pp. 79-114.

Rosenthal, R. On the social psychology of the psychological experiment: the experimenter's hypothesis as unintended determinant of experimental results. American Scientist, 1963, 51, pp. 268-283.

Rothe, H. Does higher pay bring higher productivity? Personnel, 1960, 37, pp. 20-38.

Rubeck, P. The role of the Hawthorne effect in the measurement of change. Unpublished doctoral study, Indiana University, 1971.

Runkle, Philip J. "Some recent ideas in research methodology." University of Oregon, Eugene, 1967. ED 010 221.

Sarason, S.B. The Culture of the School and the Problem of Change. Allyn and Bacon, Inc. Boston, 1971.

Saretsky, G. The consequences of an innovation as determinants of control group behavior: an exploratory study. Indiana University, 1972a, mimeo.

Saretsky, G. The OEO P.C. experiment and the John Henry Effect, Phi Delta Kappan, May, 1972, pp. 579-581.

Schon, D. Technology and Change. Dell Publishing. New York, 1967.

Shulman, L.S. Reconstruction of educational research. Review of Educational Research, 1970, 40, pp. 371-398.

Stewart, M. Resistance to technological change in industry. Human Organization, 1957, 16, pp. 36-39.

Stufflebeam, D.L. The Use of Experimental Design in Educational Evaluation. Journal of Educational Measurement, 1971a, 8, 267-274. (b)

Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., and Provus, M.M. Educational Evaluation and Decision Making. F.E. Peacock Publishers, Inc. Itasca, Illinois, 1971.

Suchman, E.A. Evaluative Research, Principles and Practices in Public Service and Social Action Programs. Russell Sage Foundation, New York, 1967.

Suppes, P. Computer-assisted instruction, Science, 1969, pp. 166, 343-350.

Suppes, P. Effectiveness of alternative instructional media. Review of Educational Research, 1974, 44, pp. 1-68.

Watson, G. Resistance to change. In G. Zaltman (Ed.). Processes and Phenomena of Social Change. Wiley Interscience. New York, 1973.

Webb, E.J., Campbell, D.T., Schwartz, R.D. and Sechrest, L. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Rand McNally & Company, Chicago, 1966.

Zaltman, G., Duncan, R., and Holbeck, J. Innovations in Organizations. Wiley Interscience. New York, in press.

Zdep, S.M., and Irvine, S.H. A reverse Hawthorne Effect in educational evaluation. Journal of School Psychology, 1970, pp. 8, 2, 89-95.